Published 2003

Page 1

Visualizing Online Auctions

Galit S HMUELI and Wolfgang J ANK

Online auctions have been the subject of many empirical research efforts in the fields of economicsandinformationsystems.Theseresearcheffortsareoftenbasedonanalyzingdata from Web sites such as eBay.com which provide public information about sequences of bids in closed auctions, typically in the form of tables on HTML pages. The existing literature on online auctions focuses on tools like summary statistics and more formal statistical methods such as regression models. However, there is a clear void in this growing body of literature in developing appropriate visualization tools. This is quite surprising, given that the sheer amount of data that can be found on sites such as eBay.com is overwhelming and can often not be displayed informatively using standard statistical graphics. In this article we introduce graphical methods for visualizing online **auction** data in ways that are informative and relevant to the types of research questions that are of interest. We start by using *profile plots* that reveal aspects of an **auction** such as bid values, bidding intensity, and bidder strategies. We then introduce the concept of *statistical zooming (STAT-zoom)* which can scale up to be used for visualizing large amounts of auctions. STAT-zoom adds the capability of looking at data summaries at various time scales interactively. Finally, we develop ***auction*** *calendars* and ***auction*** *scene* visualizations for viewing a set of many concurrent auctions. The different visualization methods are demonstrated using data on multiple auctions collected from eBay.com.

Key Words: Bid data; eBay.com; Profile plots; STAT-zoom.

1. INTRODUCTION

Almost every Internet user today has heard, browsed, or used the online **auction** site eBay.com, a major online marketplace and currently the biggest consumer-to-consumer (C2C) online **auction** place. The fascination with eBay has been documented in many recent reports and newspaper articles (*The New York Times* 2004; *USA Today* 2003). eBay has been one of the few survivors of the late 1990s electronic commerce boom. In fact, eBay has not only survived but is growing faster than ever. This has led to a surge of empirical

Galit Shmueli is Assistant Professor, and Wolfgang Jank is Assistant Professor, Department of Decision and Information Technologies, Robert H. Smith School of Business, University of Maryland, College Park, MD 20742 (E-mail addresses: gshmueli@rhsmith.umd.edu *and* wjank@rhsmith.umd.edu).

Journal of Computational and Graphical Statistics, Volume 14, Number 2, Pages 1–21
DOI: 10.1198/106186005X48236

work based on data from eBay.com, typically by researchers from the fields of economics and information systems. The issues investigated in these articles range from exploring factors that affect final prices (Lucking-Reiley, Bryan, Prasad, and Reeves 2000), analyzing the eBay reputation and feedback system (Dellarocas 2001; Livingston 2002; Resnick and Zeckhauser2002),findingempiricalevidenceforlatebidding(sniping)(RothandOckenfels 2002; Ockenfels and Roth 2002), learning about commonly encountered effects such as the "Winner's curse" (Bajari and Hortacsu 2003), detecting collusion (Kauffman and Wood 2005), investigating bidding strategies (Bapna, Goes, and Gupta 2003; Ockenfels and Roth 2002), modeling the bidder arrival process (Shmueli, Russo, and Jank 2004; Vakrat and Seidman, 2000), and more. Similar questions have been addressed by using data from other online **auction** houses such as ubid.com, amazon.com, and onsale.com. This article focuses on displaying data from eBay.com, but the methods could be adjusted for use with other

online **auction** data.

eBay offers a vast amount of rich data. Besides the time and the amount of each bid placed in each **auction**, eBay also records plenty of information about the bidders, the seller, and the product being auctioned. On any given day, several million auctions take place on eBay and all closed auctions from the last 30 days are publicly available on eBay's Web site. This huge amount of information can be quite overwhelming and confusing for the user (here we refer to the user as either the seller, a potential buyer, or the **auction** house) who wants to incorporate this information into his/her decision-making process. And of course for researchers who collect these data, it is also hard to sift through the information without appropriately visualizing it first. Although standard statistical tools like summary measures and regression models are used frequently to answer specific research questions, there is a surprising void in methods that visualize the flood of information prevalent on eBay. The lack in adequate graphical displays starts at the very beginning, in describing the raw bid data. The few articles that do attempt to use graphical displays (e.g., Lucking-Reiley et al. 2000) tend to use over-simplified plots which in some cases even distort the information contained in the data. In this article we make use of existing graphical displays as well as modify and develop new ones to visualize the information contained in bid data. Visualizations of historical auctions are useful as an exploratory tool for learning about bidding, selling, and winning on eBay.com or, more generally, in auctions. Our first aim is to expose and describe this unique type of data, which has not attracted much attention from statisticians. We point out the special features of online **auction** data and point out why ordinary statistical visualization methods require modification in some cases, while in other cases entirely new methods are needed. Our second aim is to highlight the need for adequate visualizations in the exploration of online data, and to introduce such graphics, for the first time, into the field of online **auction** research.

Raw eBay data come in the form of "bid histories," which are, from a technical point of view, HTML pages containing tables. These HTML pages are hard to grasp intuitively or to study directly, especially when looking at a multitude of concurrent auctions. Section 2 introduces typical examples of bid histories, their special structure and features, and the modern mechanisms that are used to collect them. The aim of the various visualizations

that we propose in the following sections is to assist in the explorative phase that precedes more formal analyses for answering scientific questions. Some questions of interest that would benefit from these visualizations are the amount of variation in bidding histories for similar items, bidder strategies, fraud detection (where sellers receive negative ratings for transactions), seasonal price changes, and so on.

Section 3 introduces a variety of simple visualization tools. We start by creating *profile plots*, a simple visualization of single or several bid histories, which preserves temporal information. We show what type of information is revealed by such displays and discuss their advantage over looking at the raw HTML pages. Several variations of the profile plot are illustrated, where additional features and enhancements can be integrated for various purposes of study (e.g., for exploring bidder behavior or bidding intensity throughout and **auction**). Finally, we discuss the problem of scaling profile plots for visualizing a multitude of auctions. This motivates the concept of *statistical zooming* (STAT-zoom), which we introduce in Section 4. The idea is to view data summaries at different time scales, thereby adding the capability of capturing the information contained in multiple bidding histories at a spectrum of time scales. Incorporating interactivity into the plots is known to be effective in increasing visual scalability (Eick and Karr 2002). We illustrate the STAT-zoom concept for visualizing a large number of auctions. Section 5 discusses more complex types of visualizations that are useful for visualizing multiple concurrent auctions, either for a single itemorforavarietyofitems.Twousefulvisualizationsare*calendarsofauctions*and*auction scene* maps. Section 6 concludes this article with future research directions.

2. THE DATA: BID HISTORIES ON EBAY.COM

Understanding eBay's **auction** mechanism is central to understanding the special features and structure of eBay bid data. Another important factor is the special data collection mechanism which is typically used for gathering eBay data. Here we give a brief description of the **auction** and collection mechanisms and then explain and illustrate the structure of a bid-history for a closed-ended **auction**.

2.1 T HE E BAY .COM A UCTION M ECHANISM

Most of the auctions on eBay are second-price auctions, where the highest bidder wins the **auction** and pays the second highest bid. eBay uses a proxy-bidding system where biddersaresupposedtoplacethehighestamountthattheyare*willingtopay*fortheauctioned item. These values are usually abbreviated as WTP values (Bapna et al. 2003; Roth and Ockenfels2002).Thesystemthenautomaticallyincreaseseachbidder'sbidbytheminimum increment (which is relative to the current high bid and set by eBay) until either the bidder's maximum has been reached or the bidder has the current high bid (Linoff and Berry 2001). This guarantees that the winner will pay the minimum between his/her WTP value and an increment above the second highest bid. A bidder is free to place as many bids as he/she wishes. eBay uses closed-ended auctions, where the duration of the **auction** is fixed and

predeterminedbytheseller.Open-endedor"Going-Going-Gone"auctions(suchasauctions on amazon.com) have a varying duration. In open-ended auctions the **auction** closes only after a certain amount of time passes from the last placed bid.

Duringtheongoingauctionthebidders'WTPvaluesaredisclosedexceptforthehighest bid at any moment, along with the usernames of participating bidders and the times that the bids were placed. Once an **auction** closes, eBay reveals the WTP values of all bidders except the winner. For further details on eBay's proxy system see http://pages.ebay.com/help/buyerguide/bidding-prxy.html.

A typical eBay closed **auction** page contains the sequence of bids, the bidder usernames and their feedback scores, and the exact time and date when each bid was placed. [Each user has a feedback score which reflects their experience selling and buying on eBay. When an **auction** is completed and the **transaction** is carried out, the seller and buyer have a chance to rate each other. They can assign a positive (+1), neutral (0), or negative (−1) rating. The overall rating is the sum of all ratings that a user received. For further details see http://pages.ebay.com/help/feedback/feedback-scores.html.] There is also additional information about the seller (ID and rating), the product, shipping costs, and so on. In this work we use the term "bid-history" mainly to describe the sequence of WTP values and the times they were placed.

Figure 1 displays a single closed **auction** page for a Palm M515 Personal Digital Assistant (PDA). Notice that the order in which eBay displays the bids is ascending in the WTP values, not chronologically! This makes it seem, at first, that the process of bidding was much more gradual and with higher intensity of bidding than actually occurred.

2.2 D ATA C OLLECTION A GENTS

Modern technologies allow for a convenient collection of large amounts of high quality data from the Internet. The use of Web agents or Web spiders facilitate the creation of large databases of bidding data. A Web agent is a software application, typically based on a programming language like Pearl or Java, that "crawls" over an Internet site or a collection of Web pages and gathers the desired information. In this form, data on hundreds

collection of Web pages and gathers the desired information. In this form, data on hundreds, thousands, and even more auctions can be collected in a matter of only minutes. This modern automated collection system is much less error-prone than traditional data collection and recording. Unless the data on the Web site are erroneous or not sufficiently structured, the agent will usually deliver error-free data. However, preprocessing that relies on domain knowledgeisstillneeded.Forexample,althoughmostauctionsarecarriedoutinU.S.dollars (USD), occasionally a different currency is used. Kauffman and Wood (2003) described the revolutionary aspect of new data collection mechanisms such as software agents and discuss their impact on empirical research.

Figure 1. Bid History from eBay.com for closed auction of a Palm M515 PDA.

2.3 S AMPLE D ATASET

In this article we use two sets of eBay bid data. The first dataset consists of nearly 500 closed auctions for Palm M515 personal digital assistants (PDAs) that took place between March 14 and May 25 of 2003. In an effort to reduce as many external sources of variability as possible, we collected only auctions in U.S. dollars, for completely *new* (not used) items with no added features, and where the seller did not set a secret reserve price. Furthermore, we limited the data to competitive auctions, where there were at least two bids. A subset of these bid histories (n = 158), which consist of only the seven-day auctions, are publicly available at http://www.smith.umd.edu/ceme/statistics/. The data include the bid times, bid amounts, and additional bidder information.

The second dataset consists of nearly 11,000 closed auctions for a broad variety of items that took place between August 2001 and February 2002. All auctions were in U.S. dollars, and were competitive (had at least two bids). For further information on these data see Borle, Boatwright, and Kadane (2005).

Figure 2. Profile Plots for two Palm M515 PDA auctions (the left plot corresponds to the auction in Figure 1).

Finally, the bid history for the single five-day Palm M515 PDA **auction**, which is used in the next section, appears in its complete raw form in Figure 1.

3. DISPLAYING RAW BID HISTORIES

This section looks at the raw data through informative, clear glasses. We start by displaying single auctions and then proceed to visualizing the information contained in multiple auctions.

3.1 PROFILE PLOTS: DISPLAYING SINGLE BID HISTORIES

A profile plot is a time plot of the WTP values over the duration of the **auction**. It is the first step in clarifying the information contained in a bid history. Looking at the data chronologically shows that WTP values do not affect the current level of the price, since they do not exceed the highest WTP value at that time. Figure 2 displays profile plots for two five-day auctions for a Palm M515 PDA. The plot on the left describes the **auction** from Figure 1. For this **auction**, it can be seen that after the bid of $175.25 was submitted on day 2, three lower bids followed (of $159, $169.55, and $175). The reason for this is the second-price nature, where the WTP of $175.25 is not displayed during the live **auction** until it is exceeded. Figure 2 also gives information about the intensity of bidding over time. For example, in the right panel we see an **auction** that had very little or hardly any activity at the beginning of the **auction**, followed by very intense bidding toward the end of the **auction**. In comparison, the bidding activity for the **auction** in the left panel had a very strong start, then a spurt of bids on day 2, and a final spurt at the end of the **auction**.

We can integrate more information into the profile plot, depending on the research question at hand. For example, we may be interested in the final price as a function of the values that bidders saw during the live **auction** (rather than the WTP values). Alternatively, we could be interested in the relation between the WTP values and the values that were seen during the **auction**. Due to the second price auctions that eBay uses, the WTP values are undisclosed during the **auction** until they are exceeded, and therefore they can (and are

Figure 3. Profile plot and shaded plot for a single Palm M515 PDA five-day ***auction****. Left panel: WTP values (stars) and live-bid values (step-function line) for a single* ***auction****. The horizontal line at $200.50 displays the closing price. Right panel: The area between the live bid value and the current highest WTP value.*

very likely to) be different from the current price displayed in the live **auction**, which we call *live bid values*.

To learn about the relation between the WTP values and the live bid values, we reconstruct the live **auction** bid values from the bid history by using a function that is based on the principles of the proxy-bidding mechanism and the increment rules that eBay uses (http://pages.ebay.com/help/buy/bid-increments.html). The left panel in Figure 3 displays both types of values (WTP and live bids) and the closing price for the same Palm M515 PDA described in Figures 1 and 2. The step-function describing the live bid values is always below the WTP values. This follows eBay's guarantee not to pay more than an increment above the highest bid. The graph shows the immediate effect of the $175.25 bid, of increasing the live bid value by an increment over the second highest WTP (from $81.05 to $152.5). However, because the bidders participating in the **auction** saw only the value of $152.5, it explains the arrival of the next three lower bids of $159, $169.55, and $175. The right panel of Figure 3 displays the difference between the live bid values and the current highest WTP value. This is useful for studying the ongoing "surplus" in an **auction**, which is the difference between the highest WTP and the price. Such a plot shows how fast the current price catches up with the maximum proxy bid. It can be seen, for instance, that the WTP value of $150 placed on day 2 (following the previous WTP of $80.05) creates a large area that lasts nearly 12 hours until a higher WTP value (of $159) is placed. A current

limitation of this plot for eBay data is that it does not display the surplus at the **auction** end, because at this time eBay discloses all the WTP values except for the winner's (which is the highest WTP value in the entire **auction**).

3.2 PROFILE PLOT VARIATIONS: INTEGRATING ADDITIONAL INFORMATION

Wecanusecolorandotherfeaturestoincorporateadditionalinformationintotheprofile plot. The type of information to be incorporated depends on the research question at hand.

*Figure 4. Profile plot for Palm M515 PDA, with different symbols representing different bidders. Squares represent bids of users who placed single bids. In this **auction** there are two bidders (circles and asterisks) who placed multiple bids.*

For example, researchers have been interested in examining bidding strategies. Various

authors have observed that the number of bidders on eBay is usually much smaller than the number of bids placed (Bajari and Hortacsu 2003), that is, a few bidders submit multiple bids on the same item within the same **auction**. This indicates that the bids placed at each time point are not the true willingness-to-pay values, since otherwise a bidder would not have revised his/her bid over and over again! It is therefore important to be able to visualize the behavior of different bidders, by being able to identify bids that belong to the same bidder. One option is to use different colors and/or shapes to denote different bidders. This is illustrated in Figure 4, in which an **auction** profile of the same Palm PDA is plotted with the addition of different symbols. Squares are used to represent single-time bids where the user did not place any other bids. In this **auction** there were eight bidders, with two of them placing multiple bids (represented by circles and triangles in the plot). These two persistent bidders placed 9 of the 17 bids. It is interesting to notice that one of the persistent bidders placed bids only at the very beginning (triangles), while the other (circles) seems to have monitored the **auction** and placed bids throughout its duration. The winning bid came from a single-time bidder. These three types of bidding behaviors have been reported in online auctions research and are often classified as *evaluators, participators,* and *opportunists* (Bapna et al. 2003). A useful addition to the bidder-specific profile plot is to integrate additional statistics on the prominent bidders. This can be implemented through a legend or by hovering over a point that corresponds to that bidder. The additional information can be taken from the same bid history, such as the bidder rating or ID. A more complicated task is to extract information on the bidder from a relational database that includes other auctions that this bidder participated in. An example of a useful statistic of this sort would

Figure 5. Profile plot for Palm M515 PDA, with circle size representing bidder feedback: Bigger circles represent higher feedback. The different colors represent distinct bidders (hollow circles are bids by single time bidders).

be the proportion of winnings from all the auctions that the user participated in.

Another useful variation is to use color and/or size on a profile plot to code user feedback. Bajari and Hortacsu (2003) found that experts tend to bid late in the **auction** relative to nonexperts. Furthermore, Ockenfels and Roth (2002) posited that experienced bidders will tend to place only a single bid during the last minute of the **auction**. eBay bid histories also include the feedback for users which are typically used as a measure of expertise. If this rating indeed measures expertise, then we would expect to see bids towards the end of the **auction** coming from bidders with high feedback, and those bids will tend to be single bids. On Figure 5 we use circle size to represent the bidder feedback for each bid submitted, and use hollow circles to denote bids by single-time bidders. If we disregard multiple bids by the same bidder (hollow circles), this plot shows when high-rated bidders place bids relative to low-rated bidders. It can be seen that in this **auction** the two persistent biddershaveverylowfeedback,whereasthehigherrated(moreexperienced)bidderstended to place single bids.

In conclusion, the profile plot is easily adaptable to different research questions. With some imagination, many factors of interest (e.g., day of week) can be integrated into it without clutter.

3.3 PROFILE PLOTS FOR MULTIPLE AUCTIONS

Next, we integrate the information from multiple auctions for the same item. We look at bid histories from 10 auctions for the Palm M515 PDA, each lasting seven days (this is a subset of our first dataset). Figure 6 combines the bids from the 10 auctions. The left panel displays the profile plots of the 10 bid histories. In this graph we eliminated the step

Figure 6. Profile plot (left) and shaded plot (right) for ten seven-day Palm M515 auctions (α-transparency = .2)

function for the live-bid values to reduce clutter. A graph of this type reveals several useful pieces of information about the **auction**:

- The intensity of bidding changes over time: There are two dense clusters of WTP values at the beginning (days 0–1) and especially at the end (day 7), while the middle of the auctions experiences much lower bidding activity.
- The closing prices (denoted by horizontal red lines) vary between $230–$280, with $280 being exceptionally high.
- Many WTP values were placed above the closing prices of other items. This means that the valuation for this item is highly variable: there are many people who are willing to pay substantially higher prices than others!

Another plot that can be used for exploring the "current **auction** surplus" in a set of bid histories is a shaded plot that displays the area between the WTP and live-bid values (like Figure 3, right). In order to be able to compare the overlaid auctions, we use alpha transparency. The left panel of Figure 6 illustrates this type of plot for the same 10 Palm auctions.Thedarkerareasrepresentconcentrationsofauctionswith"surplus,"andthewidth of these areas represent the time it takes the **auction** price to catch up with the WTP value.

Profile plots are useful for displaying several auctions, but they do not scale up well. Especially when conditions such as beginning price and length of **auction** vary, the profile plot becomes too cluttered and it is hard or impossible to track single auctions on it. Figure 7 illustrates this point by plotting the profiles of 158 seven-day auctions for the Palm PDAs from our first dataset. The left panel is a profile plot, while the right is a shaded plot (comparing WTP and live bid values). Note the added clutter due to varying starting prices. However, some interesting characteristics of these auctions can still be seen even on the profile plot:

profile plot:

- Activity levels are much higher on the last day of the auctions.
- The closing prices (denoted by horizontal lines) vary between approximately $175–$280 with the majority of auctions closing at around $230.

Figure 7. Profile plot (left) and shaded plot (right) for 158 seven-day Palm M515 auctions (α-transparency = .2)

- There are a few bids placed before the last day, which exceed the closing prices of some of the auctions. These appear within the horizontal line area at the top.

The shaded plot shows us that overall large surpluses are concentrated at the beginning and end of the auctions. The two black areas stretch out to approximately 12 hours, indicating the amount of time until the **auction** price catches up with the WTP values. In comparison, in the middle of the **auction** "surplus" levels and duration are much more variable from **auction** to **auction**, as reflected by multiple shades of varying time-widths. Finally, as expected, the large surplus on day 6 vanishes towards the **auction** end. This vanishing effect is artificial, because we do not have the actual highest WTP value, as explained earlier.

If the auctions have different duration, then the profile plot is even less appealing for displaying many auctions, unless the time scale is standardized. From our experience

for displaying many auctions, unless the time scale is standardized. From our experience, profile plots are useful for describing a single or several (< 30) auctions. Their usefulness is enhanced greatly by plotting auctions that have similar starting prices, that have the same duration (e.g., seven days), and that take place in a short time period of each other. In general, any factor that is known to affect the profile should be used to separate auctions into separate profile plots.

4. SUMMARIZING BID-HISTORIES

In order to learn about the characteristics of bid profiles for a certain item, bidders would ideally make use of historical data on closed auctions of the item of interest. Browsing through eBay's "bid history" pages one **auction** at a time can be overwhelming (since there are several million auctions taking place on eBay every day), and it is also hard to absorb the information on one single page due to the special structure of the HTML pages. Moreover, aside from an abundance of data, information is organized in a misleading way, since it is sorted by WTP values rather than chronological order.

Web tools that are aimed at supporting bidders' efforts, such as Andale.com or Hammertap.com, supply the user with aggregated information on historic (closed) auctions from

eBay.com. They typically give the average selling price and the number of bids. In other words, they aggregate WTP values over time and over auctions. From graphs such as Figure 4, which display the entire WTP profile for multiple auctions, it is clear that important information is lost by such aggregation. On the other hand, as the number of auctions increases and the number of bids per **auction** increases, looking at the entire individual bid profiles (of both the WTP values and live bid values), might also be overwhelming.

The question is how to summarize the entire information on multiple auctions for a certain item without losing valuable information. Instead of aggregating bid values of an entire **auction**, we suggest to aggregate over certain time-periods within the **auction** so that these time intervals are affected by the bidding intensity during different periods of the **auction**. This intensity-dependent aggregation is described next.

4.1 AGGREGATING BIDS INTERACTIVELY

From empirical research on online auction data it is known that the bid intensity changes throughout the duration of auctions. Terms such as "last-minute bidding" or "sniping" (Roth and Ockenfels 2002; Bajari and Hortacsu 2003) describe the phenomenon that towards the end of closed-ended online auctions there tends to be high bidding activity. In contrast, bidding is usually sparse during the middle of the **auction**, while bidding intensity at the start of an auction appears to vary across different items (Jank and Shmueli 2005). Shmueli et al. (2004) developed a three-phase parametric model for the bid arrival process and showed that it can capture the bid arrival process at eBay well. Thus, an optimal time-aggregation would take into account bidding intensity, such that intense periods would be aggregated only over very short periods and less-intense periods would be aggregated over longer time periods. Because we are aggregating over multiple auctions for the same item, we rely on the user's visual ability to account for the bidding intensity in the following way: In order to find a good balance between over- and under-aggregation in time, we suggest *STAT-zoom*, a **hierarchical** interactive aggregation approach. This approach is more statistically advanced than techniques suggested in the context of interactivity. It has the flavor of automatic selection aggregation (Eick 2000), but it is used for continuous data rather than categorical data. In automatic aggregation, statistics are automatically recalculated for a selection of the data chosen by the user. The selection is typically a category (e.g., unmarried females). Thus, choosing a selection of a bar chart will automatically give the statistics for the chosen selection. In our case the time scale is continuous and we treat it as a hierarchy of categories. For example, the first hierarchy could be days, then within days we have hours, then minutes, and so on. The idea is not just to show, but also to *actively compute summary statistics and/or display plots* at different time scales. Figure 8 describes this: The top panel displays daily boxplots of the bid values. STAT-zooming-in to the last two days is achieved by clicking on the last two boxplots and selecting hourly intervals. This would instantly yield the plots in the middle panel. We can further STAT-zoom-in by clicking on a boxplot of interest and obtain immediate summarizations for the interval and time scale of interest. For instance, the last two hours are plotted in the bottom panel. The depth of STAT-zooming in and out

is limited only by the units of the data. Practically, this means that we can STAT-zoom-in during periods of high activity and generate statistics and plots of the bids at frequent time intervals. During quiet periods with low activity we STAT-zoom-out, and compute averages and boxplots based on longer intervals. These graphs were created using Trellis displays (Cleveland, Shyu, and Becker 1996), implemented using the package "Lattice" in R. Separate panels are used to distinguish days in the hourly display (middle) and to distinguish hours in the minutely display (bottom). For summarizing the bid data we chose boxplots, which have the advantage of preserving many features of the bid distribution. It can be seen, for example, that the hourly bid distribution described in Figure 8 (middle) is sometimes very skewed, and thus plotting the mean or variance alone would not reveal the outliers that are of special interest in this context.

The main idea behind the STAT-zoom approach is that aggregating data at fine time resolutions will be redundant in times of low bidding activity, while aggregating at coarse time resolutions will lead to information loss during times of intensive bidding activity.

A method that is similar to STAT-zoom would be to group the data into equal-size subgroups (i.e., the intervals are chosen so that the number of observations in each interval is equal), and compute the statistic/graph for each of the subgroups. This means that during low bidding activity subgroups would include bids over longer time intervals compared to high bidding activity areas. The only manipulation with this method would be to decide on the desired subgroup size. The main advantages of STAT-zoom over equal-size subgrouping are:(1)InSTAT-zoomtheuserchoosessubgroupsoftimeintervalsthataremeaningfulinthe domain of application (such as days, or minutes), and (2) From a design and interpretation point of view, equal-size subgrouping will yield statistics/plots that are not equally spaced, whereas in STAT-zoom the intervals within a zoom level are always equal.

4.2 DISPLAYING BID INTENSITY

Although the time-aggregating boxplots account for the bidding intensity when aggregating the WTP values over time, they do not present the information on the bid intensity, that is, the amount of bidding over time. The conventional way of handling this from a statistical point of view (i.e., to describe the distribution of interest, taking into account the sample sizes), is to use boxplots with a width proportional to $\sqrt{n}$ where n is the number of aggregated bids in that boxplot (McGill, Tukey, and Larsen 1978). This method has two disadvantages in this case: First, it is useful more for the sake of comparing the boxplots (wider ones are based on more bids than narrower ones), but not for learning about the actual number of bids, which is of interest here. Second, since the display might include many boxplots when refining to fine time intervals such as minutes, varying-width boxplots would cause more clutter than reveal information. We thus suggest a different enhancement to the boxplots that allows the user to browse the WTP values and bid intensity simultaneously: we add an intensity histogram at the bottom of the graph, with the bins selected to match the aggregation level used in the boxplots. The histogram can include two vertical scales to display the counts and the percentage or cumulative count/percentage. The boxplots then

Figure 8. Illustration of STAT-zoom, by aggregating bids from 158 Palm M515 auctions at three time scales: Daily boxplots of WTP values (top), hourly boxplots for the last two days (middle), and minutely boxplots for the last two hours (bottom).

Figure 9. Daily (left) and hourly (for the last two days, right) WTP-value distribution and intensity over time for 158 Palm M515 PDA auctions.

describe the aggregated WTP value distributions and the histogram below them reveals the

number of bids in that time period. An example of a combined plot for the 158 seven-day Palm M515 PDA auctions is given in Figure 9. This was also created using Trellis displays. Separate panes (left plot) help distinguish the two days. Here we can see that the boxplots of bids during days 2–5 are based on approximately the same amounts of bids, whereas the days 1 and 6 have slightly more bids, and day 7 is based on almost four times the amount of bids. Combining the boxplot and intensity information we see that even after controlling for the amount of bids placed on that day the amount of outliers on day 7 is still surprising, and might be indicative of a mixture of two distributions.

5. VISUALIZING CONCURRENT AUCTIONS

Much insight can be gained from looking at concurrent auctions for the same item. Although most of the research on online **auction** is based on multiple auctions for the same item (or several items), only few consider the time concurrency of the different auctions in their database. For example, Zeithammer (2003) investigated the effect of the availability of multiple open auctions for the item of interest on bidding strategy and final price. Kauffman andWood(2005)examinedthepossibilityofcollusionthroughtheexaminationofamassive dataset of concurrent auctions selling the same item. As before, we have not encountered any attempts at visualizing data from this perspective.

We suggest to start looking at concurrent auctions by creating a *calendar of auctions*. This is a visualization that displays each **auction** as a line that extends between its opening and closing times. On such a graph it is possible to display auctions of various durations (e.g., eBay's 3, 5, 7, and 10 day auctions). Longer auctions are represented by longer lines. We can use different colors for different **auction** lengths. The second axis can be used for incorporating another factor of interest such as final price. Figure 10 displays an **auction** calendar for 476 auctions for the Palm M515 PDAs (all the auctions in our first dataset), where the vertical axis displays the closing price of the **auction**. The thick line represents

the daily median closing price (which is computed from the closing prices of all auctions that close on a certain day) and the shaded area extends between the daily upper and lower

that close on a certain day), and the shaded area extends between the daily upper and lower quartiles. The bottom graph depicts the daily volume of open/active auctions. We learn the following from this visualization: First, we see the period over which the data were collected, extending from mid-March through May of 2003. Around May 1 there is period of several days with a substantial decrease in **auction** activity (less than 20 open auctions). Further investigation revealed that this decrease is due to the data collector's spring vacation, and is unrelated to eBay! Other noticeable patterns are a decrease in the median closing

*Figure 10. **Auction** calendar for 476 Palm M515 PDA auctions. In the top graph, each **auction** is a line extending from the **auction** start to the **auction** end. The vertical axis is the closing price. The thick line is the median daily closing price, and the shaded area extends between the lower and upper quartiles. The bottom graph shows daily **auction** volume: the number of open/active Palm auctions in the database, by date.*

Figure 11. The eBay Auction Scene for 10,078 auctions. Grayscale represents seller rating (black = low (negative), white = high), size represents number of auctions.

price of Palm PDAs from March to May, and daily fluctuations in daily closing prices. In general, peaks in **auction** volume and in median closing prices could be related to seasonal effects such as holidays. This example illustrates the ability of the **auction** calendar to highlight interesting patterns in the data, to reveal information about the data collection, and to serve as a basis for further exploration. Second, the **auction** calendar gives a sense of how many auctions were taking place on a certain day/period. Other time related effects such as weekday/weekend effects can be examined directly from the **auction** calendar without the need to aggregate the data. For example, Lucking-Reiley et al. (2000) used a bar chart to describe the volume of **auction** closings by day-of-week. They found that more auctions tend to close on weekends relative to weekdays. Variations such as using color to represent **auction** length, weekend/weekday, or other classes in the data can therefore be useful for

visualizing the effect of different factors.

Our second suggested visualization for concurrent auctions captures a snapshot of all the auctions in a certain time period. We call it ***the auction scene***. The display is based on the **hierarchical** nature of the **auction** market, which is broken down to categories, subcategories, and so on down to the item level. The visualization uses TreeMap, a space-constrained visualization of **hierarchical** structures designed by Shneiderman in the 1990s (Shneiderman 1992; Bederson, Shneiderman, and Wattenberg 2002). TreeMap enables users to compare nodes and subtrees even at varying depth in the tree, and help them spot patterns and exceptions. Treemaps are interactive and allow dynamic querying. An electronic markets

XX,

Figure 12. The eBay Auction Scene for 10,078 auctions. Grayscale represents number of bids per auction (black = low, white = high), size represents number of distinct bidders in auction

application of TreeMap is the "Honeycomb" toolkit, developed by the Hive Group (http: //www.hivegroup.com/amazon dyn.html). It uses TreeMap to display consumer goods sold on Amazon.com.

Figure 11 displays the eBay **auction** scene for a sample of nearly 11,000 auctions that took place between August 2001 and February 2002. For further information on the data see Borle et al. (2005). The display is divided into rectangles representing categories of auctioned items (e.g., jewelry and watches). Each rectangle is then further divided into subcategories(e.g.,premiumwristwatches),andfinallyintobrands(e.g.,RolexandCartier). We can use color, size, and labels to display three variables of interest. In the figure we use a grayscale to denote seller rating (determined from feedback on previous transactions), where black denotes very low/negative rating and white very high/positive rating, and size represents the number of auctions. It is immediately apparent that very low rated sellers are concentrated almost exclusively in the premium wristwatches **subcategory**. In comparison, high-rated sellers are most common in the Dell 17-inch monitors and Oakley sunglasses items. This is of special interest because negative seller rating can be an indication of fraud. If we take into account item values, it becomes more clear why low rated sellers are concentrated in premium wristwatches: Rolex watches are sold at approximately $2,000, compared to other items in this sample that typically sell for less than $100. It is probably worthwhile for a seller to take a risk of conducting a fraudulent **auction** for a $2,000 watch but not for a $50 monitor. Figure 12 explores the relation between the number of different

XX

bidders in an **auction** and the total number of bids in an **auction** (in the eBay system a bidder can place more than a single bid). Grayscale represents the number of bids (black represents few bids, white represents many bids) and size represents the number of distinct bidders. It can be seen that although the largest number of distinct bidders is in the sports category (and especially for golf bags), the busiest items in terms of number of bids are Oakley sunglasses and Rolex wristwatches. A plausible reason for this is that golf bags are items

of broad interest, but there is no incentive to pay more than their market value. Premium wristwatches, on the other hand, appeal to a population of bidders that is considerable smaller, but who may have a stronger interest in winning the prestigious item. Furthermore, premium watches are substantially more expensive, and therefore the price increase process is "long enough" (in the sense of bid increases) for bidders to revise their bids.

The **auction** scene maps are therefore very useful for exploring the many factors that can be measured in online **auction** data. They can help detect not only relations, but also outliers and unexpected patterns. Moreover, they offer a bird's-eye view of the **auction** scene, and thus deliver an image with is usually unavailable via standard statistical displays.

6. FUTURE RESEARCH DIRECTIONS

The visualizations described in this article are meant for displaying data that have already been collected and stored. Such historic data are usually used for learning about a variety of different phenomena like bidding strategies and a seller's trustworthiness. One of the next steps is to observe and process the data in real-time. This is similar to the two phases used in control charts (in statistical quality control), where historic data are used for constructing the limits on the charts and then charts with these limits incorporated are used for monitoring real-time data. Several of the visualizations that we suggested can be used for real-time visualizations with little or no change: An **auction** profile can be used for monitoring an ongoing **auction** as long as the incoming WTP values are available. In eBay, for example, the bid history discloses the WTP value only after it has been exceeded. However,bymonitoringtheauctionusinganagent,thelivebidscanberecordedandplotted. Because the **auction** duration is known at the **auction** start, the horizontal axis can be set accordingly. An example of a slight modification would be the calendar of auctions. In a calendar that gets updated in real time we must show the right censoring somehow. One option is to mark an ongoing **auction** with a right arrow which extends to the current date. Methods based on STAT-zooming require more significant modification. Finally, real time data and their availability also call for new visualizations that would directly target their structure and the goals of monitoring them.

Withrespecttoimplementationoftheproposedvisualizations,mostcanbeeasilycoded by using standard software. We generated all the graphs with Matlab and R. However, to achieve the real-time interactivity needed for STAT-zoom, a more advanced application is needed. A further complication is that the application should be able to input data with its special structure (namely, a set of unequally spaced time series of difference duration). The software package Spotfire (www.spotfire.com) is a tool that can handle the data structure

and has many interactive options such as zooming and panning. However, since the concept of STAT-zooming is new, we have not found applications that implement it. This means that moving from one time scale to another requires, in the least, rebinning of the bid values and computing the summary statistics or graphs for the new bins. An implementation of STAT-zoom is therefore expected to be innovative.

ACKNOWLEDGMENTS

This research was partially funded by the NSF grant DMI-0205489. We thank the two anonymous referees for their invaluable suggestions, which enhanced several of our visualizations and led to the creation of others (e.g., the shaded profile plots). We also thank Ben Shneiderman for his useful comments and for suggesting the use of Treemaps for visualizing online auctions. We are grateful to the people who helped us obtain eBay data: Sharad Borle for the data used in the auction scene map; Corey Angst, Jason Kuruzovich, and Boaz Shmueli for the Palm M515 PDA data.

[Received February 2004. Revised June 2004.]

REFERENCES

Bajari, P., and Hortacsu, A. (2003), "Winner's Curse,Reserve Prices and Endogenous Entry: Empirical Insights from eBay," *RAND Journal of Economics*, 34, 329–355.

XX

Bapna R., Goes, P., and Gupta, A. (2003), "Analysis and Design of Business-to-Consumer Online Auctions," *Management Science*, 49, 85–101.

Bederson, B., Shneiderman, B., and Wattenberg, M. (2002), "Ordered and Quantum Treemaps: Making Effective Use of 2D Space to Display Hierarchies," *ACM Transactions on Graphics*, 21, 833–854.

Borle, S., Boatwright, P., and Kadane, J. B. (2005), "The Timing of Bid Placement and Extent of Multiple Bidding: An Empirical Investigation Using eBay Online Auctions," working paper, Tepper School of Business, Carnegie Mellon University, available at http://littlehurt.gsia.cmu.edu/gsiadoc/WP/2005-E27.pdf.

Dellarocas, C. (2001), "Analyzing the Economic Efficiency of eBay-like Online Reputation Reporting Mechanisms," in *Proceedings of the 3rd ACM Conference on Electronic Commerce*, 171–179.

Cleveland W. S., Shyu, M. J., and Becker, R. A. (1996), "The Visual Design and Control of Trellis Display," *Journal of Computational and Graphical Statistics*, 5, 123–155.

Eick, S. G. (2000), "Visual Discovery and Analysis," *IEEE Transactions on Visualization and Computer Graphics*, 6, 44–58.

Eick, S. G., and Karr, A. F. (2002), "Visual Scalability," *Journal of Computational and Graphical Statistics*, 11, 22–43.

Jank, W., and Shmueli, G. (2005), "Profiling Price Dynamics in Online Auctions Using Curve Clustering," working paper, Smith School of Business, University of Maryland. Available at http://www.smith.umd.edu/ceme/statistics/AuctionProfiling-Rev1-Jank&Shmueli.pdf..

Kauffman, R. J., and Wood, C. A. (2005), "The Effects of Shilling On Final Bid Prices in Online Auctions," *Electronic Commerce Research and Applications*, 4, 18–31.

(2003), "Revolutionary Research Strategies for E-Business Management: A Philosophy of Science Perspective for Research Design and Data Collection in the Age of the Internet," MIS Research Center Working Paper #03-32, Carlson School of Management, University of Minnesota.

Linoff, G. S., and Berry, M. J. A. (2001), *Mining the Web*, New York: Wiley.

Livingston, J. (2002), "How Valuable Is A Good Reputation? A Sample Selection Model Of Internet Auctions," Workshop in Industrial Organization and Microeconomic Theory.

Lucking-Reiley, D., Bryan, D., Prasad, N., and Reeves, D. (2000), "Pennies from eBay: the Determinants of Price in Online Auctions," Working paper, University of Arizona. Available at http://www.vanderbilt.edu/econ/reiley/papers/PenniesFromEBay.pdf

McGill, R., Tukey, J. W., and Larsen, W. A. (1978), "Variations of Box Plots," *The American Statistician*, 38, 12–16.

The New York Times, January 22, 2004, "Ebay's 4th Quarter Results Surpass Even its Own Hopes."

Ockenfels, A., and Roth, A. E. (2002), "The Timing of Bids in Internet Auctions: Market Design, Bidder Behavior, and Artificial Agents," Working paper #2002-33, Max Planck Institute for Research into Economic Systems, Strategic Interaction Group.

Resnick, P., and Zeckhauser, R. (2002), "Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System," *The Economics of the Internet and E-Commerce*, ed. M. R. Baye, Volume 11 of *Advances in Applied Microeconomics*, Amsterdam: Elsevier Science.

Roth, A. E., and Ockenfels, A. (2002), "Last-Minute Bidding and the Rules for Ending Second-Price Auctions: Evidence from eBay and Amazon Auctions on the Internet," *The American Economic Review*, 92, 1093–1103.

Shmueli, G., Russo, R. P., and Jank, W. (2004), "Modeling Bid Arrivals in Online Auctions," working paper, Smith School of Business, University of Maryland. Available at http://www.rhsmith.umd.edu/ceme/statistics/BidArrivals Jan04.pdf.

Shneiderman, B. (1992), "Tree Visualization with Tree-Maps: A 2-Dimensional Space Filling Approach," *ACM Transactions on Graphics*, 11, 92–99.

USA Today, December 29, 2003, "The Economy According to eBay."

Vakrat, Y., and Seidmann, A. (2000), " Implications of the Bidders' Arrival Process on the Design of Online Auctions," in *Proceedings of the 33rd Hawaii International Conference on System Sciences*, 6, 1–10.

Zeithammer, R. (2003), "Forward-Looking Bidders in Sequential Auctions: A Theory of Bargain-hunting on eBay," working paper, University of Chicago, Graduate School of Business. Available at http://ssrn.com/abstract=468847.

XX